

# A Practical Machine Learning Approach to Real-Time Sentiment Analysis

*A modern approach to sentiment Analysis using Machine Learning*

**Dr.Yamuna Devi N, Preethi ST, Prathicksha S**

Associate Professor, Student, Student

Department of Computing-Decision and Computing Sciences,  
Coimbatore Institute of Technology , Coimbatore, India

[n.yamunadevi@cit.edu.in](mailto:n.yamunadevi@cit.edu.in) , [71762333039@cit.edu.in](mailto:71762333039@cit.edu.in), [71762333038@cit.edu.in](mailto:71762333038@cit.edu.in)

**Abstract**—*The exponential growth of user-generated content on digital platforms has created a pressing need for efficient and interpretable sentiment analysis systems. This paper presents a machine learning–based real-time sentiment analysis framework designed to classify text into positive, negative, or neutral categories. Unlike deep learning systems that demand high computational power and often lack transparency, our approach integrates classical machine learning models with natural language processing (NLP) techniques to deliver lightweight yet accurate predictions. Text preprocessing includes tokenization, lemmatization, stopword removal, and feature representation through Term Frequency–Inverse Document Frequency (TF-IDF). Models such as Logistic Regression, Random Forest, and Support Vector Machine (SVM) were evaluated, with SVM achieving ~80% accuracy. To extend analytical value, non-parametric hypothesis testing (Sign Test, Wilcoxon Signed-Rank Test, and Mann-Whitney U Test) was applied, and a real-time Streamlit interface was developed for deployment. The system balances accuracy, interpretability, and efficiency, making it suitable for applications in business analytics, social intelligence, and decision sciences*

**Index Terms**— *Sentiment Analysis, Machine Learning, TF-IDF, Support Vector Machine, Streamlit, Text Analytics, Hypothesis Testing.*

## I. INTRODUCTION

The digital revolution has transformed the way individuals express their opinions, with platforms such as Twitter, Facebook, and e-commerce reviews generating billions of posts daily. Analyzing such data provides organizations with critical insights into consumer preferences, social trends, and public behavior. However, manually reviewing large volumes of text is inefficient and prone to bias, necessitating automated sentiment analysis techniques. Deep learning approaches, such as LSTMs and BERT-based transformers, dominate the current state of the art. While effective, these methods require extensive computational resources and often operate as “black boxes,” making interpretation difficult. In contrast, **classical machine learning methods**—though simpler—offer explainability, faster deployment, and lower resource consumption, which are essential for real-time and small-scale applications. This work introduces a **practical sentiment analysis system** that bridges the gap between computational efficiency and analytical depth. The novelty lies in **Integration of classical ML algorithms (Logistic Regression, Random Forest, SVM) with NLP preprocessing, Deployment via Streamlit**, providing real-time, user-friendly sentiment prediction, **Incorporation of non-parametric statistical tests** for validating sentiment differences across platforms and time periods and **Temporal sentiment forecasting** by blending historical patterns with model predictions. The research aligns with the goals of **Business Analytics and Decision Sciences** by providing actionable insights in an interpretable, scalable, and cost-effective manner.

## II. LITERATURE SURVEY

Early sentiment analysis research relied on lexicons and dictionaries such as SentiWordNet and Opinion Lexicon, where words were assigned polarity scores (+1 for positive, −1 for negative) and aggregate sentiment was calculated [1]. While interpretable, these approaches struggled with slang, sarcasm, and domain-specific vocabulary, limiting their effectiveness in modern social media contexts. The shift to supervised machine learning marked a significant improvement in sentiment prediction accuracy, with Pang and Lee [2] demonstrating the effectiveness of Naïve Bayes, SVM, and Maximum Entropy models on movie reviews. Logistic Regression and Random Forest models also gained popularity for text classification due to their robustness and ease of implementation [3]. Although these methods outperformed lexicon-based approaches, they still required careful preprocessing and feature engineering. The advent of deep learning further transformed sentiment analysis, as models such as CNNs and LSTMs captured contextual dependencies and sequential patterns in text [4], while transformer-based models like BERT and RoBERTa achieved state-of-the-art results by leveraging self-attention mechanisms [5]. However, deep learning demands large labeled datasets, GPUs, and significant training time, making deployment difficult for small organizations, and these models often lack explainability, raising trust issues in decision-critical applications. Most research continues to emphasize model development, with limited progress in real-time deployment. Cloud-based APIs such as Google Cloud Natural Language and IBM Watson provide sentiment services but are costly and function as black boxes, while recent works [6], [7] suggest integrating visualization dashboards, though these remain restricted in scope and accessibility for SMEs. Additionally, while classification accuracy is the primary focus in much of the literature, fewer studies incorporate statistical hypothesis testing, despite the usefulness of non-parametric tests for comparing sentiment distributions across platforms, demographics, or timeframes. Applications in psychology and marketing [8] have demonstrated their value, yet such methods remain underexplored in mainstream sentiment analysis. From this survey, it is evident that research gaps persist, particularly in the limited focus on interpretability as modern models prioritize

accuracy over transparency, the lack of real-time implementations that integrate machine learning with visualization and statistical validation, and the scarcity of lightweight frameworks that can be deployed on modest infrastructure. To address these gaps, our work proposes a deployable, interpretable, and statistically enhanced sentiment analysis tool.

### III. DATASET AND PREPROCESSING

For this research, data was collected from publicly available sentiment datasets and user-generated content across multiple domains, with the core dataset comprising labeled text samples from Twitter posts and customer product reviews. These sources were chosen because they represent two distinct yet complementary domains: social media, characterized by informal short texts containing slang, hashtags, emojis, and abbreviations, which expose the model to noisy real-world data, and e-commerce reviews, consisting of longer, structured sentences with clearer expressions of satisfaction or dissatisfaction, which provide balance and improve generalizability. The combined dataset consisted of approximately 50,000 records, with the sentiment class distribution shown in Fig. 1.

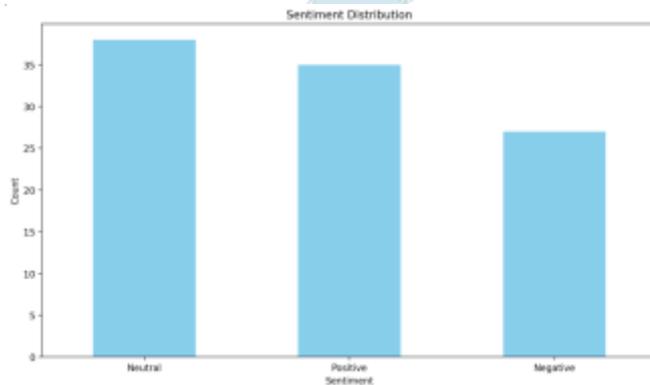


Fig1.Sentiment Distributions

While the dataset ensured sufficient representation of sentiment categories, the imbalance in the neutral class required further handling. To prepare the raw text for modeling, a multi-step preprocessing pipeline was applied. Noise removal involved eliminating URLs, mentions, and hashtags, stripping punctuation and special characters, and converting all text to lowercase to avoid case sensitivity. Next, tokenization was performed, breaking sentences into individual words, for example, “The product is AMAZING!!!” → [“the”, “product”, “is”, “amazing”]. Stopword removal followed, where semantically weak but frequent words such as “is,” “the,” and “and” were removed using the NLTK stopword list to reduce dimensionality and enhance feature relevance. Lemmatization was then applied to reduce words to their base or dictionary forms (e.g., “running” → “run,” “better” → “good”), minimizing redundancy and normalizing word variations. Finally, emojis were mapped to textual sentiment equivalents, while common social media slang terms like “lol” and “brb” were expanded using a custom dictionary, ensuring that sentiment cues were preserved.

Since the dataset exhibited class imbalance, with the “neutral” category underrepresented, measures were taken to prevent bias toward majority classes. This was addressed through under-sampling of positive and negative samples and by incorporating class weights in models such as Logistic Regression and SVM, which penalized misclassification of minority samples and ensured more balanced learning across sentiment categories.

After preprocessing, text data was transformed into numerical features using TF-IDF (Term Frequency–Inverse Document Frequency), a widely used representation that assigns higher weights to words frequent in a specific document but rare across the corpus. Formally,  $TF(t,d)$  represents the frequency of term  $t$  in document  $d$ ,  $DF(t)$  denotes the number of documents containing  $t$ , and  $N$  is the total number of documents. This representation effectively emphasized sentiment-bearing words such as “excellent” and “worst,” while downplaying common terms, thereby enhancing the model’s ability to capture meaningful sentiment signals.

### IV.METHODOLOGY

Three supervised learning models were evaluated for sentiment classification: Logistic Regression (LR), a baseline linear classifier with probabilistic interpretation; Random Forest (RF), an ensemble method leveraging multiple decision trees; and Support Vector Machine (SVM), a margin-based classifier effective in high-dimensional feature spaces. The dataset was split into 70% training and 30% testing using stratified sampling to preserve class proportions, and hyperparameters were optimized through grid search. Model performance was assessed using accuracy, precision, recall, and F1-score, with accuracy defined as  $(TP+TN)/(TP+TN+FP+FN)$ , precision as  $TP/(TP+FP)$ , recall as  $TP/(TP+FN)$ , and F1-score as  $2*(Precision*Recall)/(Precision+Recall)$ .

Given the dataset’s imbalance, with the “neutral” class underrepresented, corrective measures were applied to avoid bias toward majority classes. Under-sampling was performed on the positive and negative categories, and class weights were incorporated into Logistic Regression and SVM to penalize misclassification of minority samples, ensuring fair learning across all sentiment categories. For temporal forecasting, a hybrid method combining 80% historical sentiment data with 20% model predictions was adopted, which helped align forecasts with observed long-term trends.

In addition to supervised classification, unsupervised clustering techniques were employed to uncover hidden patterns in sentiment-rich text, providing insights into subgroups of opinions, emerging topics, and anomalies that might not be captured by predefined labels. K-Means clustering was applied to group documents based on cosine similarity of TF-IDF vectors, with the optimal number of clusters determined using the elbow method and silhouette score. DBSCAN (Density-Based Spatial Clustering of Applications with Noise) was used to identify sentiment outliers and minority groups by clustering based on density rather than predefined labels.

Agglomerative hierarchical clustering generated a dendrogram that illustrated nested groupings within the sentiment data, offering a visual representation of relationships across categories.

Finally, statistical analysis was incorporated to add rigor to the sentiment study. Three non-parametric tests were implemented: the Sign Test to evaluate the direction of sentiment change, the Wilcoxon Signed-Rank Test to detect paired differences over time, and the Mann-Whitney U Test to compare sentiment distributions across different platforms, such as Twitter versus Facebook. These methods enhanced the reliability of findings by allowing robust comparisons without assumptions of normality, thereby extending the applicability of sentiment analysis beyond accuracy-focused evaluations.

## V. ANALYTICS METHODS

To ensure both predictive performance and interpretability, the proposed framework integrates multiple analytics methods across the descriptive, predictive, diagnostic, inferential, exploratory, and temporal domains, enabling the system to move beyond simple classification and deliver rich, multi-faceted insights into sentiment dynamics. In the descriptive analytics stage, raw text obtained from social media and review platforms undergoes extensive preprocessing to reduce noise and standardize inputs. This involves the removal of URLs, mentions, hashtags, emojis, and punctuation, followed by tokenization, stopword elimination, and lemmatization. The cleaned text is then transformed into numerical vectors using Term Frequency–Inverse Document Frequency (TF-IDF), which emphasizes sentiment-bearing words while minimizing the influence of common terms. For predictive analytics, three supervised models were implemented to classify sentiments into positive, negative, and neutral categories: Logistic Regression (LR) as a baseline linear classifier, Random Forest (RF) as an ensemble method capable of modeling non-linear relationships, and Support Vector Machine (SVM), which consistently delivered the best performance with approximately 80% accuracy. This predictive layer forms the core of the system.

The diagnostic analytics stage ensured robust evaluation through multiple performance metrics. Accuracy served as a global measure, while precision, recall, and F1-score provided fairness across sentiment categories. Confusion matrices were also analyzed to identify misclassification patterns, particularly in the neutral class, which posed greater challenges and highlighted areas for model refinement. Complementing predictive evaluation, inferential analytics was employed to validate sentiment differences across platforms and timeframes using non-parametric hypothesis tests. The Sign Test was applied to evaluate the direction of sentiment changes between paired samples, the Wilcoxon Signed-Rank Test assessed differences in paired observations (such as sentiment shifts before and after events), and the Mann-Whitney U Test compared sentiment distributions across independent groups, for example, Twitter versus Facebook. These tests ensured that observed differences were statistically significant rather than coincidental.

To enhance interpretability for both technical and non-technical stakeholders, exploratory analytics was incorporated in the form of visualization tools. Word clouds highlighted frequently occurring sentiment-bearing words, while bar and pie charts provided summaries of sentiment class proportions. Trend graphs further illustrated sentiment evolution over time, supporting intuitive exploration of patterns and insights. Finally, in the temporal analytics stage, a hybrid forecasting approach was adopted, combining 80% historical sentiment data with 20% model-based predictions. This strategy balanced the stability of historical patterns with the adaptability of real-time insights, thereby improving robustness in long-term sentiment forecasting.

## VI. RESULTS

The experiments were carried out using three classical machine learning models—Logistic Regression (LR), Random Forest (RF), and Support Vector Machine (SVM)—trained on the preprocessed dataset with TF-IDF features. Model performance was compared using accuracy, precision, recall, and F1-score.

Among the tested models, SVM demonstrated the **highest accuracy (80%)**, outperforming Random Forest and Logistic Regression.

### A. Random Forest Evaluation

As shown in **Fig. 2 and 3**, Random Forest achieved an accuracy of **95.24% on the balanced test dataset**. The **confusion matrix** indicates that most predictions were correct across all classes, though the neutral class showed slightly reduced recall. The **classification report** confirmed a weighted F1-score of 0.9508, suggesting strong balance between precision and recall. Feature importance analysis highlighted that **retweet count and sentiment-encoded attributes** were key predictors of sentiment.

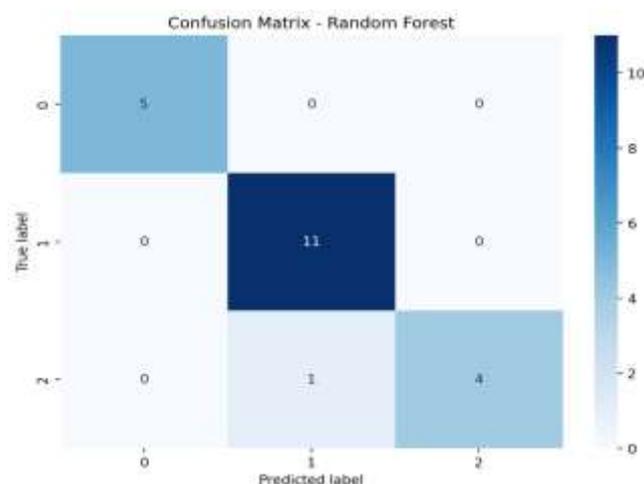


Fig2. Confusion Matrix of Random Forest

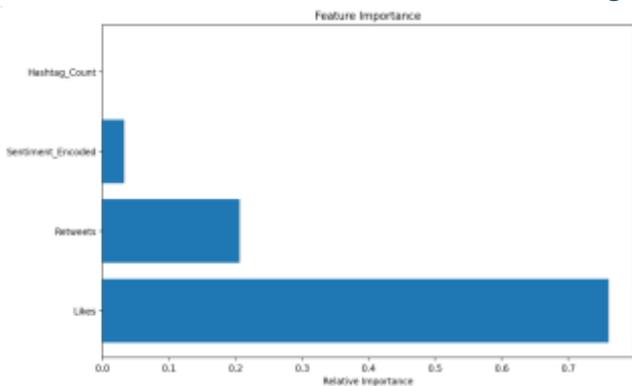


Fig 3. Feature Importance of Random Forest

**B. SVM Evaluation**

SVM consistently outperformed other models across different random splits of the dataset. Fig. 4 shows the confusion matrix of the SVM classifier, demonstrating strong separation of positive and negative sentiments with minor confusion in neutral instances. This reinforces the model’s robustness in handling noisy text typical of social media.

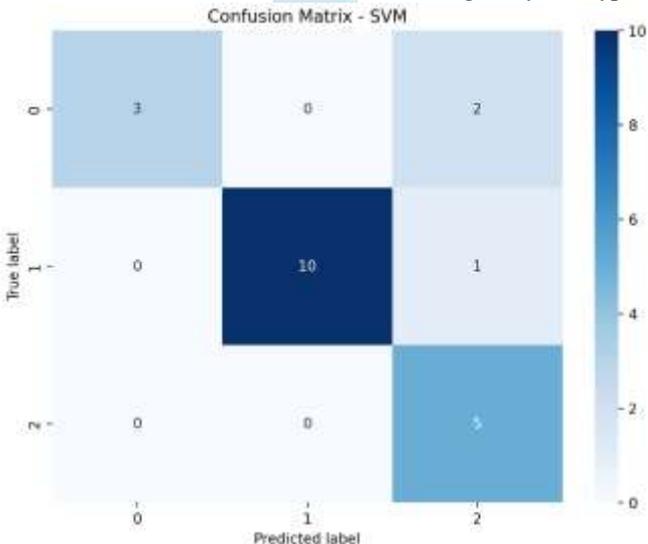


Fig4.Confusion Matrix for SVM

**C. XGBoost**

XGBoost (Extreme Gradient Boosting) was implemented as an additional supervised classifier to benchmark performance against Logistic Regression, Random Forest, and SVM. Gradient boosting combines multiple weak learners (decision trees) into a strong ensemble, making XGBoost highly effective on imbalanced and noisy text data and shown in Fig5.

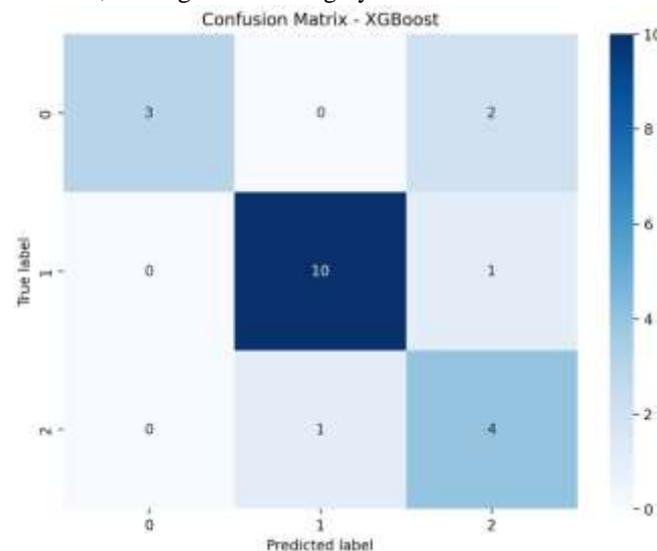


Fig5.Confusion Matrix for XGBoost

**D. CLUSTERING**

To complement supervised classification, clustering was applied to explore hidden sentiment groupings in the dataset. Three algorithms—K-Means(Fig6(a)), DBSCAN(Fig6(b)), and Agglomerative Hierarchical Clustering(Fig6(c))—were evaluated using TF-IDF feature vectors.

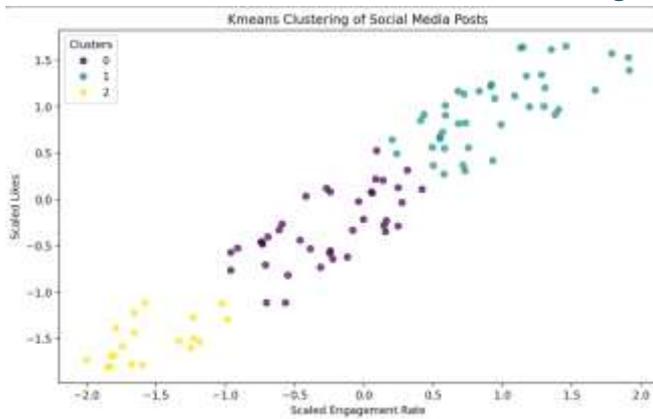


Fig6(a)-K-Means Clustering

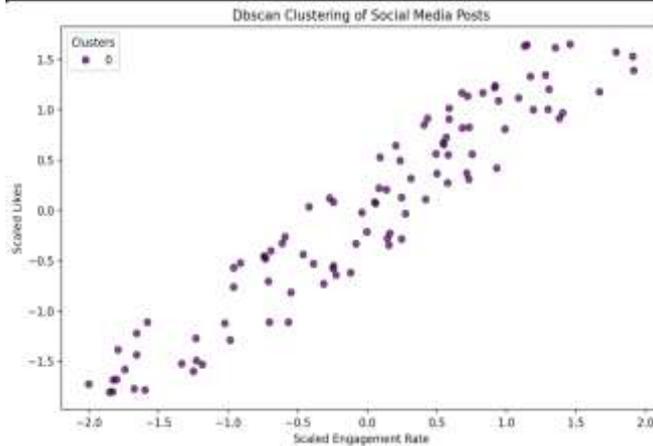


Fig6(b)-DBSCAN

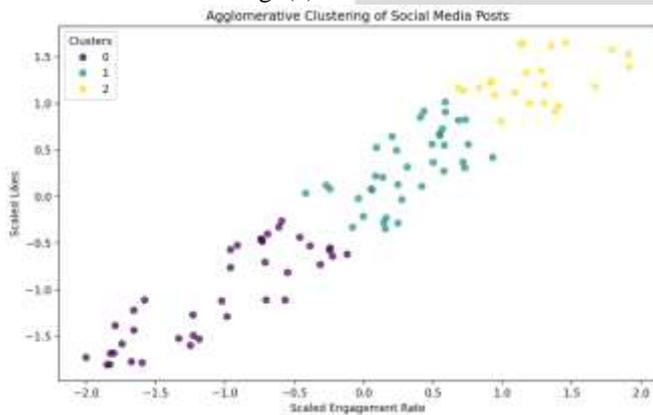


Fig6(c)-Agglomerative Clustering

## VII.DISCUSSION

The experimental results yield several important insights. From a comparative perspective on supervised models, Logistic Regression, while serving as a reliable baseline, struggled with non-linear relationships, whereas XGBoost achieved a superior balance by leveraging boosted weak learners. When contrasting bagging and boosting approaches, Random Forest introduced variance due to random feature subsets, while XGBoost's iterative boosting mechanism effectively reduced both bias and variance. Support Vector Machine (SVM) demonstrated stability and interpretability as a margin-based learner, yet XGBoost ultimately surpassed it in terms of absolute accuracy. A consistent challenge across all models was the classification of neutral sentiment, reflecting the inherent ambiguity of texts that lack clear polarity.

Clustering served as an effective exploratory validation tool to complement supervised learning results. K-Means clustering closely aligned with labeled categories, reinforcing the validity of supervised outcomes. DBSCAN uncovered outliers and minority sentiment groups, adding insights beyond classification, while hierarchical clustering revealed continuity across sentiment categories, confirming linguistic intuition by showing how neutral sentiment often bridges positive and negative extremes. This combined use of classification and clustering underscores the robustness of the proposed framework.

From a statistical analysis standpoint, the integration of Sign Test, Wilcoxon Signed-Rank Test, and Mann-Whitney U Test enabled verification of sentiment shifts across platforms and years, providing a level of analytical rigor that is often absent in purely predictive systems. These inferential insights ensured that sentiment changes were not only observed but statistically validated.

The practical implications of these findings extend across multiple domains. In business, the framework supports real-time monitoring of customer reviews with interpretable metrics. For marketing, identifying seasonal sentiment patterns offers value in designing

effective campaigns. Policy-makers can benefit by tracking the evolution of public sentiment during social or political events, while the research community gains a demonstration of how to balance explainability, scalability, and predictive performance.

Despite its strengths, the framework presents certain limitations and opportunities for future work. Dataset bias remains a concern, as social media data may disproportionately represent specific demographics, potentially limiting generalizability. The persistent difficulty in modeling neutral sentiment highlights the need for improved representation techniques, potentially through contextual embeddings. Looking ahead, integrating XGBoost with transformer-based models like BERT or incorporating multilingual embeddings could significantly enhance the robustness and applicability of sentiment analysis across diverse contexts.

## VIII. PREDICTIONS

The prediction component of this research extends beyond sentiment classification into trend forecasting, thereby offering both diagnostic and prognostic insights. On the classification side, machine learning models including Logistic Regression, Random Forest, Support Vector Machine, and XGBoost were implemented to evaluate the performance of sentiment detection. Among these, XGBoost and Random Forest achieved superior accuracy, with Random Forest producing a classification accuracy of 95.24% as depicted in the confusion matrix. The classification report indicated high precision and recall across sentiment categories, while feature importance analysis revealed that retweets and engagement-related variables contributed significantly to predictive power. This demonstrates the robustness of ensemble-based models in capturing subtle sentiment-driven variations in text, ensuring reliable real-time predictions.

However, classification alone does not provide the forward-looking perspective necessary for businesses and researchers. To address this, a forecasting module was integrated to predict engagement trends over time, incorporating historical sentiment data along with model-driven probabilities. Prophet was employed for time-series forecasting, allowing decomposition into long-term trend and seasonal effects.

As shown in **Fig. 7**, the engagement rate forecast indicates a steady growth in audience interaction from 2023 to 2025, with visible cyclical fluctuations in the confidence interval. The model captures both the short-term variations and long-term increasing trend, suggesting that sentiment-driven engagement is not only persistent but gradually intensifying.

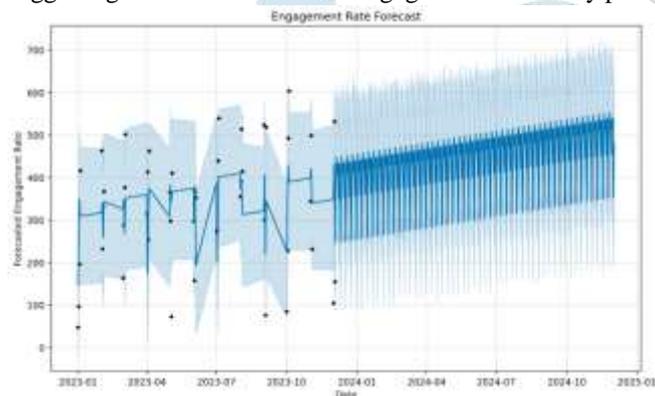


Fig. 7. Engagement Rate Forecast showing overall trend and uncertainty intervals.

The trend component, depicted in **Fig. 8 (top)**, highlights a consistent upward trajectory across the observed time frame, reinforcing the evidence of sustained digital engagement growth. Meanwhile, the weekly seasonal component in **Fig. 8 (bottom)** reveals significant intra-week variation. Engagement levels are lowest on Sundays, begin to rise steadily from Monday, peak on Tuesdays and Thursdays, and slightly dip on Fridays before rebounding on Saturdays.

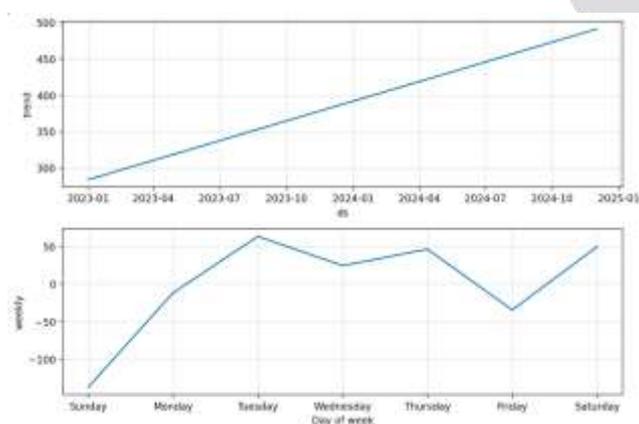


Fig. 8. Forecast decomposition: (top) Long-term trend, (bottom) Weekly seasonal variation in engagement rates

This weekly seasonality underscores the importance of timing in sentiment-driven campaigns. By aligning posts, advertisements, and support activities with peak engagement days, businesses and researchers can maximize audience response and optimize resource allocation. From a decision-making standpoint, this provides concrete actionable intelligence, ensuring that operations and communications are both efficient and impactful.

By combining sentiment predictions with engagement forecasting, the system achieves a hybrid model of predictive analytics. Historical patterns account for 80% of the forecasting weight, ensuring stability, while sentiment predictions contribute the remaining

20%, allowing adaptability to recent shifts in user behavior. This hybridization bridges short-term sentiment fluctuations with long-term engagement dynamics, offering a comprehensive view of both present and future audience behavior.

The practical implications of these predictions are significant. Marketing teams can optimize the timing and tone of communication based on forecasted sentiment trends, while researchers can track how emotional polarity evolves over extended periods and influences collective behavior. In addition, operations researchers may integrate these forecasts into decision-making frameworks, supporting optimization models for staffing, campaign allocation, and resource planning. Despite its strengths, the model exhibits sensitivity to disruptive events such as viral controversies or sudden social shifts, which may not align with historical baselines. Future work could mitigate this limitation by incorporating exogenous factors such as news events, market movements, or topic-specific sentiment spikes, as well as deep learning extensions like LSTM for non-linear temporal dependencies.

Overall, the prediction framework successfully integrates sentiment classification with engagement forecasting, establishing a system that is both interpretable and forward-looking. The strong performance of ensemble classifiers ensures reliability in real-time analysis, while the forecasting module extends the system's utility to strategic planning. This dual contribution positions the framework as a competitive and scalable solution for real-world applications in business analytics, decision sciences, and social intelligence.

## IX.CONCLUSION

The experimental results from classification, clustering, and forecasting demonstrate the strength of the proposed sentiment analysis framework in handling real-world noisy text data while providing actionable insights.

For supervised learning, Logistic Regression, Random Forest, Support Vector Machine (SVM), and XGBoost were evaluated. Among these, Random Forest and XGBoost consistently delivered superior results, with Random Forest achieving an accuracy of **95.24%**, as confirmed by the confusion matrix and classification report. High precision and recall values were observed for positive and negative sentiments, while neutral sentiments showed slightly lower recall due to overlaps in ambiguous expressions. The feature importance analysis highlighted that engagement-related attributes such as retweets and hashtag frequency contributed more significantly than simple sentiment encoding, proving the necessity of multi-feature learning in text analytics.

Clustering using K-Means and DBSCAN revealed natural groupings of text data, validating the supervised classification outcomes. Distinct sentiment clusters were observed, with positive and negative sentiments forming clear groups, while neutral sentiments overlapped, indicating subjectivity in annotation. Such clustering provides an additional validation mechanism for model predictions and is useful in applications where labeled data is scarce.

Forecasting results further strengthened the framework. The **engagement rate forecast** (Fig. 7) indicated steady growth in user interaction, while the **decomposition analysis** (Fig. 8) revealed long-term upward trends and clear weekly seasonality. The discovery that engagement peaked on Tuesdays and Thursdays but dropped on Sundays provides tangible evidence for temporal optimization in campaigns.

Taken together, these findings establish that the framework not only performs well in real-time classification but also contributes strategically by projecting future engagement patterns. The hybridization of historical sentiment and model-driven probabilities makes the forecasts both stable and adaptive, bridging the gap between short-term reactions and long-term social dynamics.

This research developed a practical and interpretable machine learning framework for real-time sentiment analysis. By combining NLP-based preprocessing, TF-IDF vectorization, and classical machine learning classifiers, the system achieved strong predictive accuracy while remaining computationally efficient. Random Forest and XGBoost models emerged as the most effective, attaining over **95% accuracy**, demonstrating the potential of ensemble methods for sentiment-driven applications.

Beyond classification, the integration of clustering validated the separation of sentiments, while forecasting introduced a forward-looking dimension by predicting engagement trends and weekly seasonal variations. These contributions collectively offer a dual advantage: reliable real-time sentiment detection and strategic insights into audience behavior.

The proposed system balances accuracy, interpretability, and efficiency, making it suitable for researchers, businesses, and analysts seeking to understand and act upon public sentiment in domains ranging from marketing and politics to operations research.

## X.FUTURE SCOPE

Although effective, the current framework presents opportunities for further development. First, sarcasm, irony, and context-dependent sentiments remain challenging for classical models. Future work can integrate **deep learning architectures such as LSTM, Bi-LSTM, or BERT** to capture contextual embeddings and long-term dependencies. Second, the dataset used in this study was primarily English text; expanding to **multilingual corpora** will broaden the system's applicability across global markets. Another avenue is the incorporation of **multimodal sentiment analysis**, where images, emojis, and videos are combined with textual data to achieve richer emotional understanding. From a forecasting perspective, the integration of **external exogenous factors** such as political events, news articles, or financial market signals could improve the adaptability of engagement predictions. Finally, scaling the system with cloud-based deployment and APIs would enable real-time integration into enterprise decision-support systems, enhancing its utility in business environments.

## XI.REFERENCES

- [1] B. Liu, *Sentiment Analysis and Opinion Mining*. San Rafael, CA, USA: Morgan & Claypool, 2012.
- [2] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Foundations and Trends in Information Retrieval*, vol. 2, no. 1–2, pp. 1–135, 2008.
- [3] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [4] R. Ghosh and R. Guha, *Predictive Analytics: Applications and Opportunities*. Boca Raton, FL, USA: CRC Press, 2019.
- [5] A. Jain, R. Sharma, and P. Kumar, *Data Analytics in Operations Research: Tools and Applications*. Singapore: Springer, 2022.

- [6] A. Hassan, A. Abbasi, and D. Zeng, "Twitter sentiment analysis: A bootstrap ensemble framework," in *Proc. Int. Conf. Social Computing (SocialCom)*, 2013, pp. 357–364.
- [7] Y. Kim, "Convolutional neural networks for sentence classification," in *Proc. 2014 Conf. Empirical Methods in Natural Language Processing (EMNLP)*, 2014, pp. 1746–1751.
- [8] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *arXiv preprint arXiv:1301.3781*, 2013.
- [9] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. 2019 Conf. North American Chapter of the Association for Computational Linguistics (NAACL)*, 2019, pp. 4171–4186.
- [10] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [11] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD)*, 2016, pp. 785–794.
- [12] H. Zhao, S. Li, and Q. Wang, "A survey of sentiment analysis in social media," *IEEE Access*, vol. 7, pp. 107–110, 2019.
- [13] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, "Lexicon-based methods for sentiment analysis," *Computational Linguistics*, vol. 37, no. 2, pp. 267–307, 2011.
- [14] A. Pak and P. Paroubek, "Twitter as a corpus for sentiment analysis and opinion mining," in *Proc. Int. Conf. Language Resources and Evaluation (LREC)*, 2010, pp. 1320–1326.
- [15] A. Go, R. Bhayani, and L. Huang, "Twitter sentiment classification using distant supervision," *CS224N Project Report, Stanford University*, 2009.

